5-1980

A Confidence Interval Estimate of Percentile

How Coung Jou
Utah State University

A CONFIDENCE INTERVAL ESTIMATE OF PERCENTILE

by

Jou, How Coung

A thesis submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Applied Statistics

UTAH STATE UNIVERSITY
Logan, Utah

1980

ACKNOWLEDGMENTS

I sincerely appreciate Dr. Ronald V. Canfield for his assistance and criticism on this paper.

Also, I would like to extend my thanks to Dr. David L. Turner and Dr. Gregory W. Jones for their kind help.

Jou,How-Coung

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

A Confidence Interval Estimate of Percentile

by

Jou, How Coung, Master of Science

Utah State University, 1980

Major Professor: Dr. Ronald V. Canfield
Department: Applied Statistics

The confidence interval estimate of percentile and its applications were studied. The three methods of estimating a confidence interval were introduced. Some properties of order statistics were reviewed. The Monte Carlo Method-- used to estimate the confidence interval was the most important one among the three methods. The generation of ordered random variables and the estimation of parameters were discussed clearly. The comparison of the three methods showed that the Monte Carlo method would always work, but the K-S and the simplified methods would not.

(45 pages)

CHAPTER I

INTRODUCTION

Statistical analysis has become a very important part of the preliminary work in dams and dikes design. It is very important to understand the flooding characteristics of the water system which the structure serves. Thus, many design criteria include capacity to contain the "N year flood" or simulate some measure of the flow. "N year flood" is the yearly maximum of daily stream flows which is exceeded with probability 1/N, where N is specified.

The usual method of determining the N year flood is to record the yearly maximum for Y years, select a representative distribution, then estimate the parameters. The N year flood is estimated as the $(N-1)/N^{th}$ percentile of the estimated distribution. There are other methods which are also used to determine the design flood for dams and dikes.

No matter how the design flood is determined, the available information is the observed data. Therefore it is subject to the same inadequacies of any estimate of a random phenomenon. It is not precisely determined. The usual statistical characterization of this lack of the precision is the confidence interval. However for the case of design flood, no attempt has been made to estimate its accuracy. It seems that such an evaluation should be a necessity when the consequences of inadequate design are considered.

The problem of deriving confidence limits for percentiles of a distribution are considered in this thesis. An existing method using the Kolomogorov-Smirnov statistic is shown to be inadequate for the high (or low) percentiles, and a new method based on Monte Carlo simulation is proposed.

A review of the Kolomogorov-Smirnov confidence interval and of the distribution of order statistics fundamental to later derivations is given in Chapter II. The new confidence interval using Monte Carlo simulation is derived in Chapter III, and applications of this method are given in Chapter IV. A simpler method which does not involve Monte Carlo simulation is evaluated in Chapter V. This method has some intuitive appeal but as noted in Chapter V is very biased for the higher percentiles. The conclusions and recommendations of this study are summarized in Chapter VI.

CHAPTER II

REVIEW

The Kolomogrov-Smirnov confidence interval on a distribution function could be used to derive a confidence interval on the percentiles of the distributions. This method uses the sample distribution function $F_n(x)$ with sample size n.

$$F_n(x) = \frac{j}{n} \text{ for } x_j < x < x_{j+1},$$

$$j = 0; 1, \ldots, n$$

$$(x_0 = -\infty,\ x_{n+1} = +\infty).$$

This function will generally differ from the population distribution function. But if it differs from an assumed distribution F(x) by too much, we will reject the hypothesis that F(x) is the population distribution function. That is, the amount of the difference between the empirical and assumed distribution function should be a usual tool in determining whether or not to accept the assumed distribution as correct.

The least upper bound of $|F_n(x) - F(x)|$ is the statistic used to test H_0: the population distribution function is F(x). That statistic is known as the Kolomogrov-Smirnov statistic :

$$D_n = \sup_x |F_n(x) - F(x)|$$

This statistic has a known distribution under H_0. From Table 1 we can find the critical value for rejecting H_0 with a specified n and α. For large n the asymptotic values for certain α level are given in Table 1.

Table 1. Asymptotic critical values of the Kolomogrov-Smirnov method

α	0.2	0.15	0.1	0.05	0.01
Limitation	$\frac{1.07}{\sqrt{n}}$	$\frac{1.14}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.03}{\sqrt{n}}$

The statistic D_n is two-sided, involving the "absolute" difference of F(x) and $F_n(x)$. The critical region is D_n > table value. Using this property, a confidence interval with significant level α can be derived. Another method for calculating the asymptotic percentiles is from the limiting distribution:

$$\lim_{n\to\infty} p(D_n < \frac{Z}{\sqrt{n}}) = 1-2 \sum_{j=1}^{\infty} (-1)^{j-1} \exp(-2j^2 Z^2)$$

$$\simeq 1-2 \exp(-2Z^2).$$

Through upper limit and lower limit we can get a confidence interval (x_1, x_2), as shown in Figure 1.

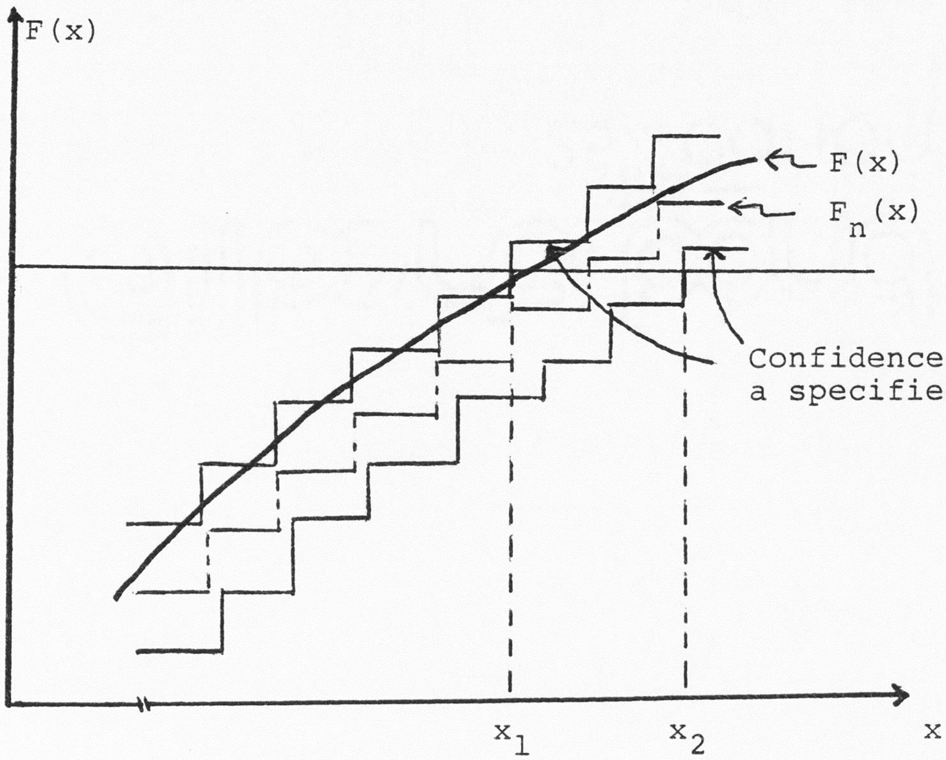

Figure 1. K-S method.

Order statistics play an important role in statistical inference partly because some of their properties do not depend on the distribution from which the random sample is obtained. Let x_1, x_2, ..., x_n denote a random sample from a distribution of the continuous type having probability density function f(x). Let y_1 be the smallest of these x_i, y_2 be the next x_i in order of magnitude, ..., and y_n the largest x_i, i.e., $y_1 < y_2$, ..., $< y_n$. Then y_i, i = 1, 2, ..., n, is called the i^{th} order statistic of the random sample x_1, x_2, ..., x_n. The density of a continuous random variable may be defined as the derivatives of the cumulative distribution function. Let $f(x)$ and $F(x)$ represent the density and cumulative distribution functions respectively of a random variable X. Then by definition of derivative

$$f(x) = \lim_{h \to 0} \frac{F(x+h) - F(x)}{h} . \qquad (1)$$

The numerator on the right of (1) can be interpreted as the probability of the event that the random variable lies in the interval $(x, x+h)$.

Consider now the k^{th} order statistic Y_k a sample of size n of the random variable of X. It will now be shown that the probability density function Y_k is

$$g_k(y_k) = \frac{n!}{(k-1)!\ (n-k)!} F(y_k)^{k-1}(1-F(y_k))^{n-k} f(y_k)$$

$$k = 1, 2, \ldots, n. \qquad (2)$$

As in (2), the probability that Y_k lies in the interval (y_k, y_k+h) will be used to derive the density function. This event requires that (a) k-1 observations lie in $(-\infty, y_k)$ (b) 1 observation lies in (y_k, y_k+h) and (c) n-k observation lie in (y_k+h, ∞). The probability that events (a), (b) and (c) occur simultaneously is

$$Ph = \frac{n!}{(k-1)!(n-k)!} F(y_k)^{k-1}(F(y_k+h) - F(y_k)) (1-F(y_k+h))^{n-k} . \qquad (3)$$

Thus

$$g_k(y_k) = \lim_{h\to 0} [\frac{P_h}{h}]$$

$$= \frac{n!}{(k-1)!(n-k)!} F(y_k)^{k-1}(1-F(y_k))^{n-k}f(y_k). \quad (4)$$

Consider now the transformed random variable U = F(X) where F(.) is the cumulative density function of X. The c.d.f. of U is

$$P(U \leq u) = P(F(X) \leq u) = P(X \leq F^{-1}(u)) = u .$$

Therefore U is uniformly distributed on [0, 1].

Let $y_1, y_2, \ldots, y_n$ be the order statistics from a sample of size n of the random variable X. Let $U_1 = F(y_1)$, $U_2 = F(y_2)$, ..., $U_n = F(y_n)$. Because F(x) is monotone increasing, the smallest U is the transform of the smallest X, and so forth, so that the ordered U's are, respectively, the transforms of the ordered X's:

$$U_k = F(y_k).$$

The distribution of $(U_1, U_2, \ldots, U_n)$ is, therefore, the distribution of the order statistics of a random sample from the uniform population on [0, 1].

Since F(x) = x for the uniform distribution, it follows from (4) that the probability density function of u_k is

$$g_{u_k}(y) = \frac{n!}{(k-1)!(n-k)!} \; y^{F-1}(1-y)^{n-k}, \; 0 \le y \le 1. \qquad (5)$$

This is a Beta distribution with parameters v = k, w = n-k+1.

CHAPTER III

MONTE CARLO CONFIDENCE INTERVAL

In this chapter a confidence interval estimate of the p^{th} percentile of a distribution is developed. The technique is based upon a method of estimation developed for the Weibull distribution (see Bain and Antls, 1968) but which can be adapted for many other distributions. This method of estimation is best explained by example. The Weibull and Normal distribution are illustrated here.

Let X be a Weibull random variable. Then

$$F(x) = 1-e^{-(x/\theta)^r}, \quad x \geq 0.$$

Let $y_1, y_2, \ldots, y_n$ and $u_1, u_2, \ldots, u_n$ be the order statistics of a sample of size n of the Weibull and an independent uniform random variable respectively. The parameters r and θ are estimated by choosing values which provide the "best fit" of F(x) through the points (y_k, u_k), $k = 1, 2, \ldots, n$. The "best fit" criterion may be least squares from $\sum_i (u_i - F(y_i))^2$. However, in practice it is convenient to transform the values $u_k = F(y_k)$ and y_k so that a simple linear relationship holds. For the Weibull case $\ln(-\ln(1-F(x)) = r \ln x - r \ln \theta$. Thus $\ln(-\ln(1-u_k))$ has a linear relationship to $\ln(y_k)$ and the least square fit is the regression line through the points $(\ln(-\ln(1-u_k))$, $\ln(y_k))$, $k = 1, 2, \ldots, n$. The slope of this line estimates r and the intercept estimates $-r\ln\theta$. Let $\eta_k = \ln(-\ln(1-u_k))$ and $r_k = \ln(y_k)$ for $k = 1, 2, \ldots, n$. Then it follows that

$$\hat{\gamma} = \frac{\Sigma\eta_k\gamma_k - \Sigma\eta_k\Sigma\gamma_k/n}{\Sigma\gamma_k^2 - (\Sigma\gamma_k)^2/n} \tag{6}$$

and

$$\hat{\theta} = \exp\left(-\frac{\Sigma\eta_k}{\hat{r}n} - \frac{\Sigma\eta_k}{n}\right). \tag{7}$$

Consider the y_k, $k = 1, 2, \ldots, n$ values fixed. For each Monte Carlo sample u_k, $k = 1, \ldots, n$, there results estimates of θ and r in turn the estimated p-th percentile x_p. The process of sampling $u_1, u_2, \ldots, u_n$ and estimating x_p is repeated for a large number of times, say 500. Then the confidence interval is interpolated from the empirical distribution function of x_p. For example, let $x_{p1}, x_{p2}, \ldots, x_{p500}$ be the ordered values of 500 x_p's computed as described previously. Then a 0.95 level confidence interval is $(x_{p12.5}, x_{p487.5})$, where $x_{p12.5} = (x_{p12} + x_{p13})/2$, $x_{p487.5} = (x_{p487} + x_{p488})/2$. Since any interval containing 475 x_{pi} values is a 95% confidence interval it is reasonable to explore the distribution to find the narrowest interval which contains 475 points. This procedure is time consuming, however, and usually differs very little from the equal tails interval given previously.

The normal distribution illustrates another method of deriving the confidence interval on x_p. Let $y_1, y_2, \ldots, y_n$ and $u_1, u_2, \ldots, u_n$ be the order statistics from random samples of size n from the normal and uniform populations. As noted previously, the u_k represent possible values of $F(y_k)$. Each sample $u_1, u_2, \ldots, u_n$ is apriori equally likely. Let $\Phi(z)$ be the standard normal distribution function (CDF). Then $\Phi^{-1}(u_k) = z_k$, $k = 1, \ldots, n$ are likely values of $(y_k-\mu)/\sigma$. Therefore there is a linear relationship between z_k and y_k where the slope is $1/\sigma$ and the intercept is $-\mu/\sigma$. Using least squares,

$$\hat{\sigma} = \left(\frac{\Sigma y_k z_k - \Sigma y_k \Sigma z_k/n}{\Sigma y_k^2 - (\Sigma y_k)^2/n}\right)^{-1} \tag{8}$$

and

$$\hat{\mu} = \hat{\sigma}\Sigma z_k/n + \Sigma y_k/n \quad . \tag{9}$$

As before the x_p is sampled by obtaining Monte Carlo samples $u_1, u_2, \ldots, u_n$. Then x_p is computed using as parameters the estimates μ and σ obtained from each Monte Carlo sample $u_1, u_2, \ldots, u_n$ with the fixed sample $y_1, y_2, \ldots, y_n$. The confidence interval is interpolated from the empirical distribution of the x_p's.

CHAPTER IV

APPLICATIONS

In this chapter the method developed in Chapter III is applied to the Weibull distribution. Generation of Weibull and uniform random variables is considered first. In the final section, the Monte Carlo confidence interval is illustrated.

Generations of Ordered Random Variables

The usual method of generating order statistics of a random variable with distribution F(x) is to generate independent uniform values, U_i, $i = 1, \ldots, n$. Then using the inverse of F(x), transform the U_i to $X_i = F^{-1}(U_i)$.

The method is more efficient if the uniform random variables are generated as order statistics. Thus avoiding the operation of ordering the sample. This is accomplished using Fortran subroutine ORDER, the method is given by Hartley and Lurie (1972) in the following subroutine.

```
SUBROUTINE ORDER (X,N,M)
DIMENSION X(N)
TEMP=0.0
SEED=TIME(11)
DO 10 I=1,M
V=RANDOM(SEED)
```

```
      U=1.0-(1.0-TEMP)*V**(1.0/(FLOAT(M-I)+1.0))
      X(I)=U
   10 TEMP=U
      RETURN
      END
```

<u>Estimation of Parameters</u>

Let y_i, i = 1, ..., n represent the order statistics of a random variable with Weibull Density Function (CDF)

$$F(x) = 1-e^{-(x/\theta)^r}, \quad x \geq 0.$$

It was shown in the previous chapter that x_i, i = 1, ..., n represent a random sample of observations of a random variable with CDF F(x). These values must then be avoided thereby greatly decreasing the cost of Monte Carlo experiments which require order statistics. In here, I do not specify the method to get those x_i, i = 1, ..., n. Replacing X(I) = U with THETA*(-ALOG(1.0-X))**(1.0/R) the inverse function of CDF of Weibull distribution. In Chapter III for the Weibull case we get $\ln(-\ln(1.0-F(x))) = r\ln x - r\ln\theta$. Thus $\ln(-\ln(1.0-U_k))$ was a linear relationship to $\ln(y_k)$ and the least square fit (where y_k is independent uniform values, k = 1, ..., n) is the regression line through the points $(\ln(y_k), \ln(-\ln(1.-u_k)))$, k = 1, ..., n. The slope of this line estimates r and intercept estimates $-r\ln\theta$.

So we can get estimates of r and θ. The Fortran program 1 listed in Appendix A generates the y_i, i = 1, ..., n and repeatedly generates uniform order statistics, with each new set of u_i, i = 1, ..., n the parameters r and θ are estimated. These estimates are then used in the program to compute the pth percentile for several values of p. In Appendix D is the data when p = 0.9, r = 2.0, θ = 10.0 and n = 40. The first column is ordered Weibull random values y_k, k = 1, ..., 40, the second column is $\ln(y_k)$, k = 1, ..., 40, the third column is ordered uniform random values u_k, k = 1, ..., n, the fourth column is $\ln(-\ln(1.0-u_k))$, k = 1, ..., n. The plot u_k vs y_k (k = 1, ..., 40) and plot $\ln(-\ln(1.0-u_k))$ vs $\ln(y_k)$ (k = 1, ..., 40) are listed in Figures 2 and 3 respectively.

The estimated parameters $\hat{r} = 1.60999$, $\hat{\theta} = 11.5163$ and $x_{0.9} = 19.332724$. The 95% confidence interval on x_p (in the thesis I try to do 500 times) has been obtained from the Monte Carlo distribution of x_p. The results are shown in Table 2.

Table 2. 95% confidence interval on the pth percentile (Monte Carlo Method)

p	Lower bound	Upper bound
0.9	12.77745	20.26675
0.95	13.68255	24.84900
0.975	15.60380	29.59265
0.99	17.02290	31.09065
0.995	19.63265	35.89530

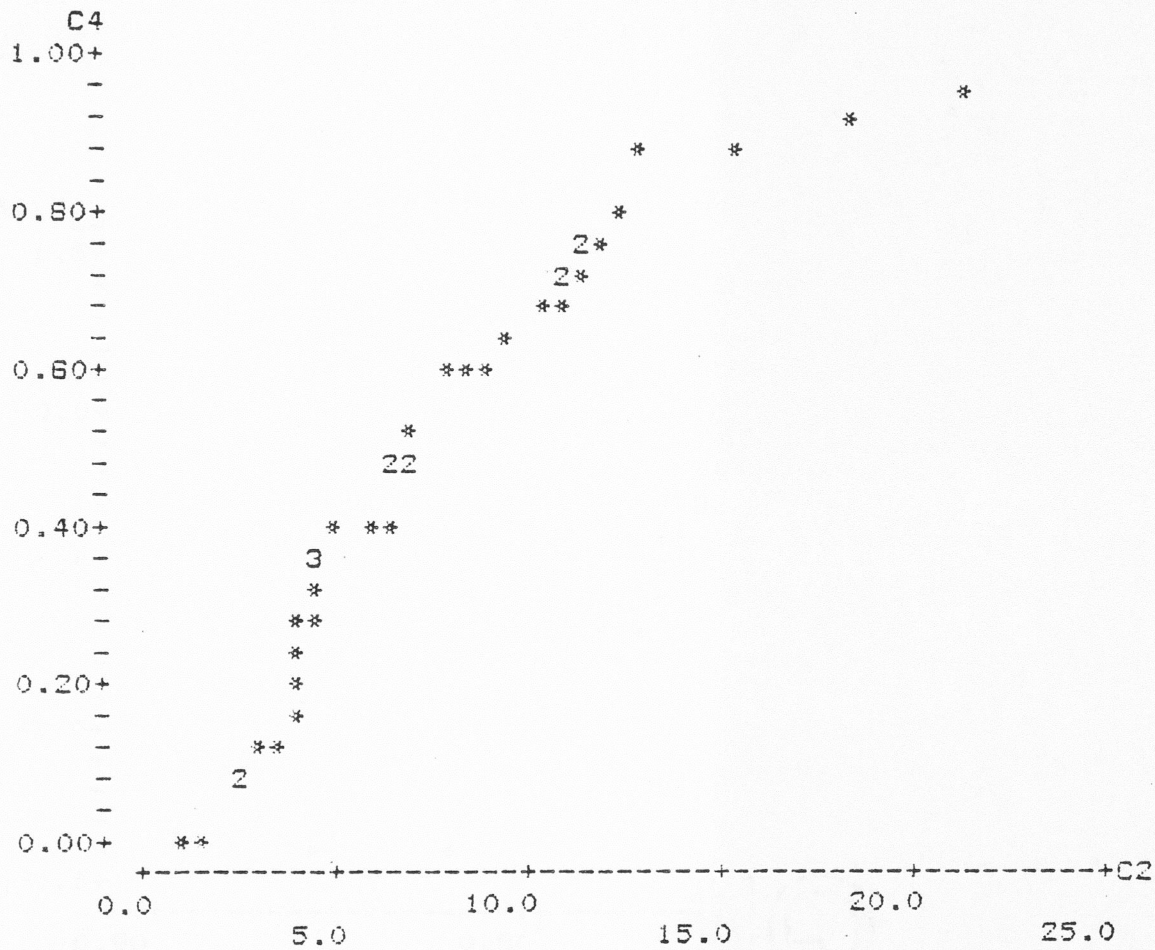

Figure 2. Plot u_k vs y_k.

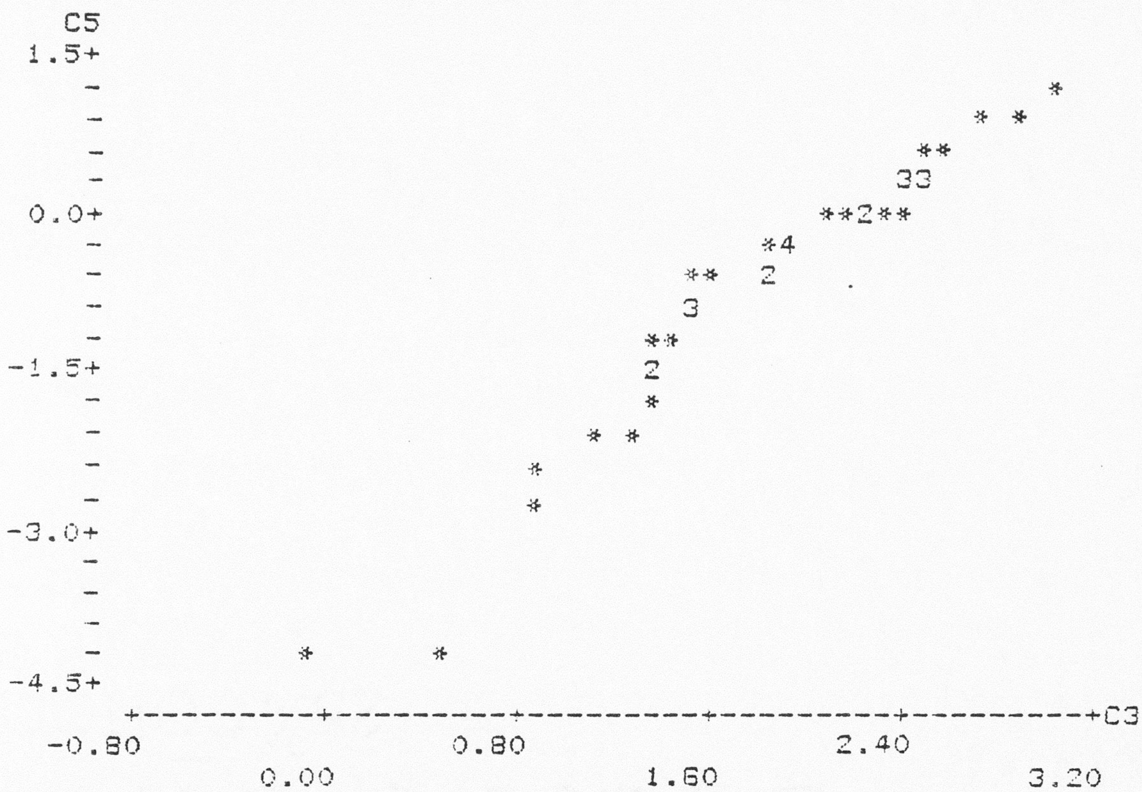

Figure 3. Plot $\ln(-\ln(1-u_k))$ vs $\ln(y_k)$.

Table 3 shows the corresponding confidence interval computed by the method of Kolomogorov-Smirnov using the same initial sample from the Weibull distribution. The Fortran program 2 listed in Appendix B is to calculate pth percentile. Note that in every case the upper bound for U_k, $k = 1, \ldots, n$ is 1.

Table 3. 95% confidence interval on the pth percentile (K-S Method)

P	Lower bound	Upper bound
0.9	12.6619730	**
0.95	13.6153340	**
0.975	12.4740910	**
0.99	13.3881968	**

** = no meanings. For example, p = 0.99, the upper bound 24.02172 has p value 0.75. It is much less than 0.99. So I say the upper bound received from exterpolating is meaningless. The data for $r = 2.0$, $\theta = 10.0$ and $n = 40$, $p = 0.95$ are shown in Appendix E.

CHAPTER V

SIMPLIFIED METHOD

The method described in Chapter III provides an approximation on the pth percentile whose accuracy is determined by the Monte Carlo sample size. Since this can be expensive if a very accurate interval is needed, a simpler method is evaluated in this section. Let $U_i = F(y_i)$, $i = 1, \ldots, n$ where y_i, $i = 1, \ldots, n$ are the order statistics of a random sample of size n from a population with CDF F(x).

It was shown in Chapter II that the distribution of U_i is Beta with parameters $v = k$, and $w = n-k+1$. It seems intuitively reasonable to construct a confidence envelope for the CDF in the following manner. At each y_i, $i = 1, \ldots, n$, construct an $(1 - \alpha)$ level confidence on U_i, $i = 1, \ldots, n$. Denote the upper and lower bounds U_i and L_i respectively. Then using the same estimation techniques used in Chapter IV, the Weibull parameters are estimated with the set of upper bound values (U_i, $i = 1, \ldots, n$) substituted for the U_i in the equation and then repeating the estimation with the set of lower bounds (L_i, $i = 1, \ldots, n$). The two resulting estimated CDF's constitute an envelope of possible CDF's based upon a $1 - \alpha$ level confidence interval.

The proposed confidence interval on the pth percentile is found by determining the percentile estimates from each of the estimated CDF's. Graphically the procedure is illustrated in Figure 4.

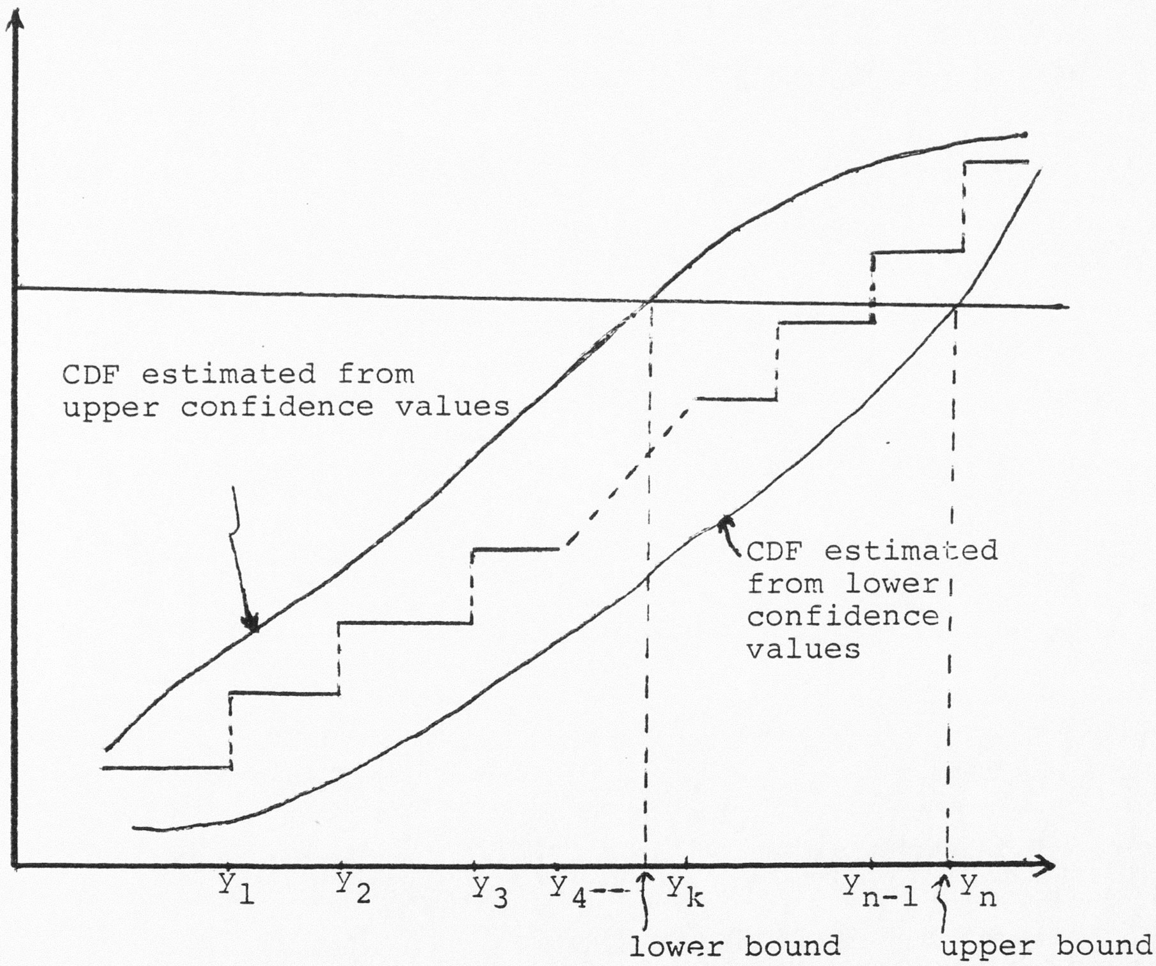

Figure 4. Confidence interval by simplified method.

The L_i and U_i, i = 1, ..., n are found as follows for 1 - α interval

$$\frac{1}{B(r,\theta)}\int_0^{L_i} t^{r-1}(1-t)^{\theta-1}\,dt = \alpha/2$$

$$\frac{1}{B(r,\theta)}\int_0^{U_i} t^{r-1}(1-t)^{\theta-1}\,dt = 1-\alpha/2. \qquad (10)$$

An approximate solution in L_i and U_i is given in the Handbook of Mathematical Function edited by M. Abramowitz and I. A. Stegun (1970) where

$$\frac{1}{B(r,\theta)}\int_0^{x_p} t^{r-1}(1-t)^{\theta-1}\,dt = p \qquad 0 \le p \le 1$$

$$x_p = \frac{r}{(r+\theta e^{2w})}, \qquad w = \frac{y_p(h+\theta)^{1/2}}{h} - (1/(2\theta-1))$$

$$- 1/(2r-1))(\lambda+5/6 - 2/(3h))$$

$$h = 2(1/(2r-1) + 1/(2\theta-1))^{-1}, \quad \lambda = \sqrt{\ln(1/p^2)}$$

$$y_p = t - \frac{c_0+c_1t+c_2t^2}{1+d_1t+d_2t^2+d_3t^3} + \varepsilon(p)$$

$$c_0 = 2.515517,\ c_1 = 0.802853,\ c_2 = 0.010328$$

$$d_1 = 1.432788,\ d_2 = 0.189269,\ d_3 = 0.001308. \qquad (11)$$

The absolute value of the error in x_p for this approximation is given as less than $4.5 * 10^{-4}$ (see program in Appendix B).

The method was applied to the same Weibull sample that was in Chapter IV, so that comparisons with the exact method could be made. The results were disappointing in that a large bias occurs in the high percentile region. The intervals are shown in Table 4 with the corresponding intervals from the exact method. The large bias in the higher percentiles is evident. The extent of the bias is seen graphically in Figure 5.

Table 4. 95% confidence interval on the pth percentile (Simplified method)

P	Lower bound	Upper bound
90%	12.702997	16.354500
95%	14.132521	17.453690
97.5%	15.162933	18.135656
99%	16.423828	18.505649
99.5%	16.961182	18.592846

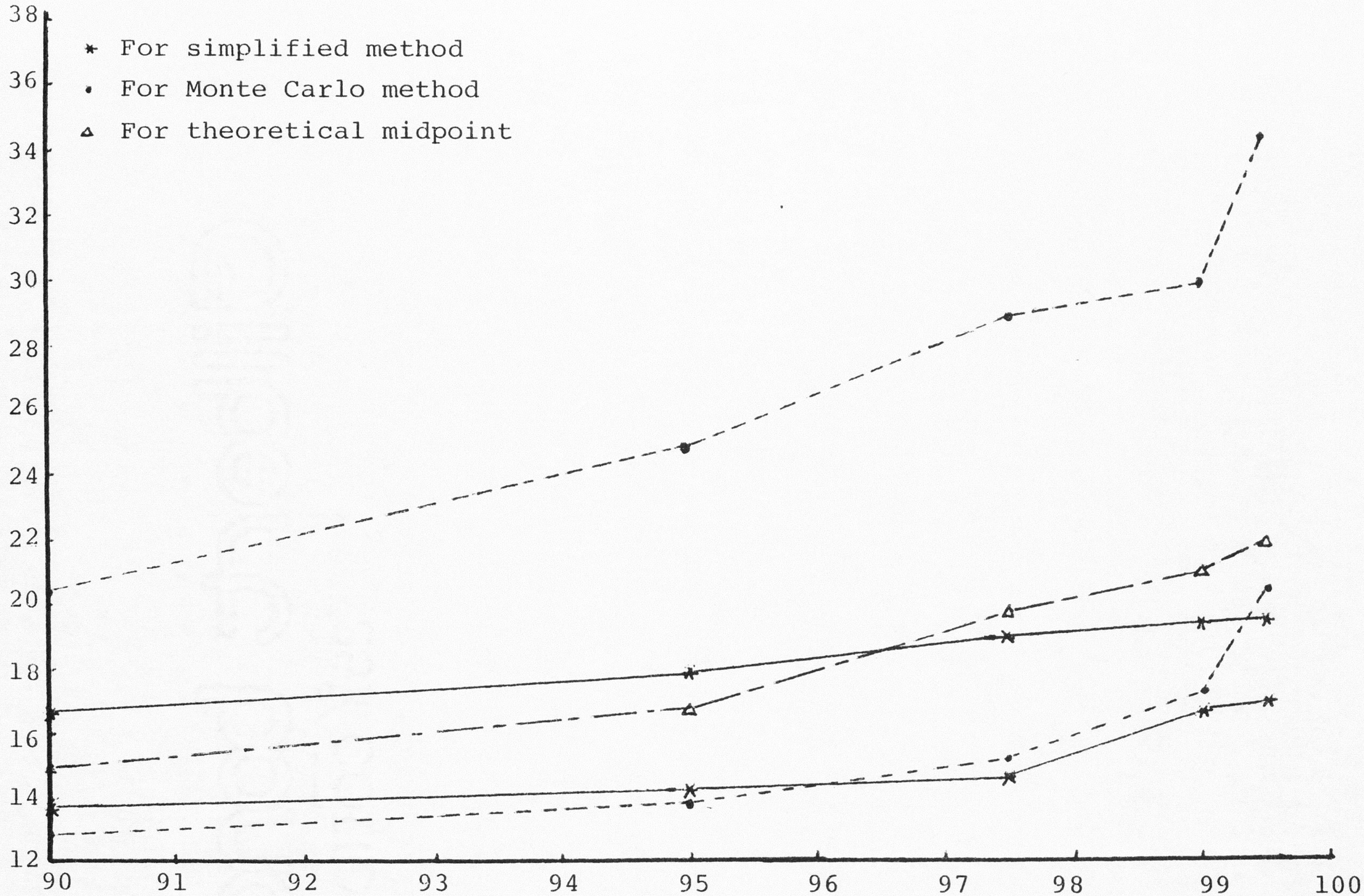

Figure 5. The plot of confidence interval vs different percentiles.

CHAPTER VI

CONCLUSIONS

Two methods for computing a $1-\alpha$ level confidence interval were developed and illustrated. The method in Chapter IV is limited in precision only by the Monte Carlo sample size. Therefore it is possible to specify any reasonable level of precision before computing the interval. The method works for any distribution which can be inverted, i.e., there exists a solution for $F(x) = U$ which can be computed. The primary disadvantage is the expense of Monte Carlo simulation.

The second method was developed on intuitive grounds and is not based on firm theoretical principles. It is evident from Figure 5 that the method does not give reasonable intervals for the high percentiles. It is instructive to compare the confidence interval with the true value of the population percentile. The Weibull data used in the computation of the intervals in Tables 2 and 4 were generated using the Weibull distribution function

$$F(x) = 1-e^{-(x/10)^2}, \; x \geq 0.$$

The true percentiles for this distribution are tabulated in Table 5.

Table 5. Percentiles of the Weibull distribution $F(x) = 1-\exp(-(x/10)^2)$

Percentile	90%	95%	97.5%	99%	99.5%
Value	15.1742	17.308*	19.2064	21.4596	23.0180

$^*17.308 = 10.0*(-\ln(1-0.95))^{1/2.0}$

A rough check on the method of computing the confidence interval is to see if the intervals cover the true value at each percentile. Note from Table 4 that the Monte Carlo method does provide intervals which cover the true value at every point. However the short method indicates a severe bias at the higher percentiles (i.e. > 90%). Thus it is clear that this method is not good.

REFERENCES

Abramowitz, M., and Stegun, I. A. 1970. Handbook of Mathematical Function. National Bureau of Standards. New York: Dover Publications.

David, H. A. 1970. Order Statistics. New York: Wiley Press.

Hartley, H. L., and Lurie, B. W. 1972. The moment of Log-Weibull Order Statistics. Technometrics, 11(2):373-385.

Hastings, N. A. J., and Peacock, J. B. 1975. Statistical Distribution. New York: Halstead Press.

Hogg, R. V., and Craig, A. T. 1972. Introduction to Mathematical Statistics. 3rd edition. New York: Macmillan Press.

Hollander, M., and Wolfe, D. A. 1973. Nonparametric Statistical Methods. New York: Wiley Press.

APPENDICES

Appendix A

Monte Carlo Method

```
#FILE (B08115)PROGRAM1 ON PACK
         DIMENSION X(50),UNIF(50),A(10,10),SUM(10),ARRAY(1001)
       DOUBLE PRECISION A,TEST,DET
       COMMON R,THETA
       READ(5,/) N,R,THETA,P
       READ(5,/) ITIME
       PP=P
       TEMP=THETA
       WRITE(6,1)
     1 FORMAT(4X,"I",10X,"ESTIMATED R",10X,"ESTIMATED THETA"///)
        CALL ORDER(X,50,N,1)
        DO 5 I=1,N
    5   X(I)=ALOG(X(I))
        DO 160 IIII=1,ITIME
C* GENERATE ORDER STATISTICS OF UNIFORM DIST. U(0,1).
        CALL ORDER(UNIF,50,N,2)
C* USE LEAST SQUARE METHOD TO FIND THE ORDER STATISTIC VALUE WHE
     \ Y=P
        DO 60 I=1,10
        SUM(I)=0.0
        DO 60 J=1,10
   60   A(I,J)=0.0
        DO 70 J=1,N
   70   SUM(1)=X(J)+SUM(1)
        DO 80 I=1,N
        UNIF(I)=ALOG((-1.)*ALOG(1.-UNIF(I)))
   80   SUM(2)=UNIF(I)+SUM(2)
        DO 90 I=1,N
        A(1,1)=X(I)*X(I)+A(1,1)
        A(2,2)=UNIF(I)*UNIF(I)+A(2,2)
   90   A(1,2)=X(I)*UNIF(I)+A(1,2)
        DO 100 I=1,2
        DO 100 J=I,2
        A(I,J)=A(I,J)-SUM(I)*SUM(J)/FLOAT(N)
  100   A(J+1,I)=A(I,J)
        TEST=0.5E-10
         CALL DMATIV(A,1,1,2,2,DET,TEST,10)
        B0=SUM(2)/FLOAT(N)
        B0=B0-A(1,2)*SUM(1)/FLOAT(N)
C* COMPUTE THE INTERSECTION POINT
        P=ALOG((-1)*ALOG(1.0-P))
        T=(P-B0)/A(1,2)
        ARRAY(IIII)=EXP(T)
        THETA=DEXP((-1)*B0/A(1,2))
        WRITE(6,122) IIII,A(1,2),THETA
  122   FORMAT(I5,E21.6,E25.6)
        P=PP
        THETA=TEMP
  160   CONTINUE
        DO 200 I=1,ITIME-1
        J=ITIME-I
```

```
      IFLAG=0
      DO 250 K=1,J
      A1=ARRAY(K)
      A2=ARRAY(K+1)
      IF(A1.LE.A2) GO TO 250
      E=A1
      ARRAY(K)=A2
      ARRAY(K+1)=E
      IFLAG=1
  250 CONTINUE
      IF(IFLAG.EQ.0) GO TO 300
 200  CONTINUE
 300  WRITE(6,303)
  303 FORMAT(////4X,"I",10X,"INTERACTION POINTS"///)
      WRITE(6,305) (I,ARRAY(I),I=1,ITIME)
  305 FORMAT(I5,E27.6)
      STOP
      END
      SUBROUTINE ORDER(X,N,M,II)
      DIMENSION X(N)
      COMMON R,THETA
      TEMP=0.0
      SEED=TIME(11)
      DO 10 I=1,M
      V=RANDOM(SEED)
      U=1.0-(1.0-TEMP)*V**(1.0/(FLOAT(M-I)+1.0))
      IF(II.EQ.1) GO TO 1
      X(I)=U
      GO TO 10
    1 X(I)=THETA*(-ALOG(1.-U))**(1./R)
   10 TEMP=U
      RETURN
      END
```

Appendix B

Simplified Method

```
#FILE (808115)PROGRAM2 ON PACK
      DIMENSION Y1(40),Y2(40),X(40),SUM(5),A(5,5)
      DOUBLE PRECISION A,DET,TEST
      COMMON R,THETA
      READ(5,/) DN,N
      READ(5,/) R,THETA
      DO 10 I=1,N
      TEMP=FLOAT(I)/FLOAT(N)
      Y1(I)=TEMP+DN
   10 Y2(I)=TEMP-DN
      CALL ORDER(X,N)
      DO 200 I=1,N
  200 WRITE(6,250) Y1(I),Y2(I),X(I)
  250 FORMAT(3X,3(F16.8,2X))
      IFLAG=1
    5 DO 15 I=1,2
      SUM(I)=0.0
      DO 15 J=I,2
   15 A(I,J)=0.0
      SUM(1)=SUM(1)+X(I)
   20 SUM(2)=SUM(2)+Y1(I)
      DO 30 I=1,N
      A(1,1)=A(1,1)+X(I)*X(I)
      A(1,2)=A(1,2)+X(I)*Y1(I)
   30 A(2,2)=Y1(I)*Y1(I)+A(2,2)
      DO 40 I=1,2
      DO 40 J=I,2
      A(I,J)=A(I,J)-SUM(I)*SUM(J)/FLOAT(N)
   40 A(J+1,I)=A(I,J)
      TEST=0.5E-10
      CALL DMATIV(A,1,1,2,2,DET,TEST,5)
      BO=SUM(2)/FLOAT(N)
      BO=BO-A(1,2)*SUM(1)/FLOAT(N)
      ANS=(0.95-BO)/A(1,2)
      WRITE(6,100) ANS
  100 FORMAT(//" THE INTERACTION PT IS",F18.7//)
      IF(IFLAG.EQ.2) GO TO 150
      DO 50 I=1,N
   50 Y1(I)=Y2(I)
      IFLAG=2
      GO TO 5
  150 STOP
      END
      SUBROUTINE ORDER(X,N)
      DIMENSION X(N)
      TEMP=0.0
      SEED=TIME(11)
      DO 10 I=1,N
      V=RANDOM(SEED)
      U=1.0-(1.0-TEMP)*V**(1.0/(FLOAT(N-I)+1.0))
      X(I)=PINV(U)
```

```
   10 TEMP=U
      RETURN
      END
      REAL FUNCTION PINV(U)
      COMMON R,THETA
      A=(-ALOG(1.-U))**(1./R)
      PINV=THETA*A
      RETURN
      END
```

Appendix C

Kolomogrov-Smirnov Method

```
308115)PROGRAM3 ON PACK
   DIMENSION X(40),Y(40),A(10,10),SUM(10)
   COMMON R,THETA
   DOUBLE PRECISION A,DET,TEST
   READ(5,/) R,THETA
   READ(5,/) N,P
   CALL ORDER(X,N)
   DO 5 I=1,N
 5 X(I)=ALOG(X(I))
   IFLAG=0
    PP=P
    SIGLEV=0.025
100 DO 8 I=1,3
    SUM(I)=0.0
    DO 8 J=1,3
  8 A(I,J)=0.0
    DO 50 I=1,N
    SUM(1)=SUM(1)+X(I)
 50 A(1,1)=A(1,1)+X(I)*X(I)
    IF(IFLAG.GE.2) GO TO 200
    CALL CINTVL(Y,N,SIGLEV)
    DO 10 I=1,N
 10 Y(I)=ALOG(-ALOG(1.-Y(I)))
    DO 12 I=1,N
    SUM(2)=SUM(2)+Y(I)
    A(1,2)=A(1,2)+X(I)*Y(I)
 12 A(2,2)=A(2,2)+Y(I)*Y(I)
    DO 15 I=1,2
    DO 15 J=I,2
    A(I,J)=A(I,J)-SUM(I)*SUM(J)/FLOAT(N)
 15 A(J+1,I)=A(I,J)
    TEST=0.5E-10
    CALL DMATIV(A,1,1,2,1,DET,TEST,10)
    B0=SUM(2)/FLOAT(N)
    B0=B0-A(1,2)*SUM(1)/FLOAT(N)
    WRITE(6,20) B0,A(1,2)
 20 FORMAT(" THE REGRESSION COEFFICIENTS B0=",F16.6,3X,"B1=",F16
  \.6//)
    P=ALOG(-ALOG(1.0-P))
    T=(P-B0)/A(1,2)
    T=EXP(T)
    WRITE(6,25) PP,T
 25 FORMAT(" THE INTERVAL BOUND FOR P=",F7.3,3X,"IS",F15.6///)
    SIGLEV=.975
    IFLAG=IFLAG+1
    P=PP
    GO TO 100
200 STOP
    END
    SUBROUTINE CINTVL(Y,N,P)
    DIMENSION Y(N)
    T=SQRT(ALOG(1./(P*P)))
```

```
      T=T-((2.515517+.802853*T+.010328*T**2.)/(1.+1.432788*T+\
    \.189269*T**
     -2.+.001308*T**3.))
      TEMP=(T**2.-3.)/6.0
      DO 10 I=1,N
      TA=1./(2.*FLOAT(I)-1.)
      TB=1./(2.*FLOAT(N-I+1)-1.)
      TH=2.*1./(TA+TB)
      W=T*SQRT(TH+TEMP)/TH
      W=W-(TEMP+5./6.-2./(3.*TH))*(TB-TA)
      Y(I)=FLOAT(I)/(FLOAT(I)+FLOAT(N-I+1)*EXP(2.*W))
   10 CONTINUE
      RETURN
      END
      SUBROUTINE ORDER(X,N)
      COMMON R,THETA
      DIMENSION X(N)
      TEMP=0.0
      SEED=TIME(11)
      DO 10 I=1,N
      V=RANDOM(SEED)
      U=1.0-(1.0-TEMP)*V**(1.0/(FLOAT(N-I)+1.0))
      X(I)=((-ALOG(1.-U))**(1./R))*THETA
   10 TEMP=U
      RETURN
      END
```

Appendix D

Part of Data for Monte Carlo Method

COLUMN	C2	C3	C4	C5
COUNT	40	40	40	40
ROW				
1	0.9359	-0.06625	0.014987	-4.19304
2	1.6086	0.47539	0.016467	-4.09808
3	2.3596	0.85851	0.075076	-2.55048
4	2.4457	0.89435	0.098586	-2.26538
5	2.9903	1.09538	0.115653	-2.09633
6	3.5200	1.25847	0.131777	-1.95682
7	3.8876	1.35778	0.140987	-1.88407
8	3.9030	1.36175	0.191368	-1.54923
9	3.9680	1.37827	0.222948	-1.37734
10	3.9929	1.38451	0.283124	-1.10005
11	4.3097	1.46087	0.285931	-1.08834
12	4.4652	1.49632	0.329663	-0.91635
13	4.4823	1.50013	0.356598	-0.81874
14	4.6720	1.54158	0.366233	-0.78510
15	4.7280	1.55349	0.379499	-0.73976
16	4.7697	1.56229	0.405183	-0.65489
17	6.2375	1.83058	0.406720	-0.64992
18	6.4346	1.86169	0.408831	-0.64311
19	6.5235	1.87542	0.475429	-0.43823
20	6.6281	1.89132	0.480704	-0.42269
21	6.8535	1.92476	0.484213	-0.41240
22	7.0346	1.95083	0.489277	-0.39760
23	7.0378	1.95130	0.539669	-0.25385
24	7.9576	2.07413	0.581934	-0.13683
25	8.4196	2.13057	0.595305	-0.10024
26	9.0841	2.20652	0.597643	-0.09385
27	9.4884	2.25007	0.656930	0.06749
28	10.4333	2.34500	0.664764	0.08885
29	10.8922	2.38805	0.685885	0.14669
30	11.0811	2.40524	0.714754	0.22666
31	11.1832	2.41441	0.724856	0.25500
32	11.4623	2.43907	0.725868	0.25785
33	11.5096	2.44318	0.742368	0.30470
34	11.6614	2.45629	0.750998	0.32952
35	11.9025	2.47675	0.769577	0.38379
36	12.2834	2.50825	0.793725	0.45650
37	12.9944	2.56452	0.867904	0.70519
38	15.6430	2.75002	0.890875	0.79537
39	18.3492	2.90958	0.904744	0.85492
40	21.4923	3.06769	0.955888	1.13816

Appendix E

Part of Data for K-S Method

0.24000000	-0.19000000	0.73087311
0.26500000	-0.16500000	2.33184649
0.29000000	-0.14000000	2.48551604
0.31500000	-0.11500000	2.61515526
0.34000000	-0.09000000	2.77417887
0.36500000	-0.06500000	2.77701459
0.39000000	-0.04000000	3.12876232
0.41500000	-0.01500000	3.54922147
0.44000000	0.01000000	3.67359461
0.46500000	0.03500000	3.74741833
0.49000000	0.06000000	4.80848285
0.51500000	0.08500000	5.06404995
0.54000000	0.11000000	5.37647384
0.56500000	0.13500000	5.96092552
0.59000000	0.16000000	6.10206654
0.61500000	0.18500000	6.30997937
0.64000000	0.21000000	6.31985085
0.66500000	0.23500000	6.53653539
0.69000000	0.26000000	6.66760516
0.71500000	0.28500000	7.06874834
0.74000000	0.31000000	7.76320753
0.76500000	0.33500000	8.29272076
0.79000000	0.36000000	8.29703542
0.81500000	0.38500000	8.34855679
0.84000000	0.41000000	8.89939393
0.86500000	0.43500000	8.93245899
0.89000000	0.46000000	9.44444534
0.91500000	0.48500000	11.65657047
0.94000000	0.51000000	11.68644308
0.96500000	0.53500000	11.75066906
0.99000000	0.56000000	12.87024329
1.01500000	0.58500000	12.92405328
1.04000000	0.61000000	13.40591624
1.06500000	0.63500000	13.66685459
1.09000000	0.66000000	14.00764642
1.11500000	0.68500000	14.16145718
1.14000000	0.71000000	14.93361613
1.16500000	0.73500000	15.62837948
1.19000000	0.76000000	16.25721020
1.21500000	0.78500000	19.00057828

The first column is for upper bound.
The second column is for lower bound.
The third column is for Weibull order random numbers
when $r = 2.0$, $\theta = 10.0$, $p = 0.95$, $n = 40$.